The Claims

1.   (Currently amended) A method comprising:

receiving a frame of content;

automatically detecting a candidate area ~~for a new face region~~ in the frame that may include a face;

using one or more hierarchical verification levels to verify whether a human face is in the candidate area;

indicating that the candidate area includes [[a]] the face if the one or more hierarchical verification levels verify that a human face is in the candidate area; and

using a plurality of cues to track each verified face in the content from frame to frame.

2.   (Original) A method as recited in claim 1, wherein the frame of content comprises a frame of video content.

3.   (Original) A method as recited in claim 1, wherein the frame of content comprises a frame of audio content.

4.   (Original) A method as recited in claim 1, wherein the frame of content comprises a frame of both video and audio content.

5.   (Original)   A method as recited in claim 1, further comprising repeating the automatic detecting in the event tracking of a verified face is lost.

6.   (Original)   A method as recited in claim 1, wherein receiving the frame of content comprises receiving a frame of video content from a video capture device local to a system implementing the method.

7.   (Original)   A method as recited in claim 1, wherein receiving the frame of content comprises receiving the frame of content from a computer readable medium accessible to a system implementing the method.

8.   (Currently amended)   A method as recited in claim 1, wherein detecting the candidate area ~~for the new face region in the frame~~ comprises:

detecting whether there is motion in the frame and, if there is motion in the frame, then performing motion-based initialization to identify one or more candidate areas;

detecting whether there is audio in the frame, and if there is audio in the frame, then performing audio-based initialization to identify one or more candidate areas; and

using, if there is neither motion nor audio in the frame, a fast face detector to identify one or more candidate areas.

9.   (Currently amended) A method ~~as recited in claim 1~~ <u>comprising:</u>

<u>receiving a frame of content;</u>

automatically detecting a candidate area for a new face region in the frame, wherein detecting the candidate area ~~for the new face region in the frame~~ comprises:

determining whether there is motion at a plurality of pixels on a plurality of lines across the frame;

generating a sum of frame differences for each possible segment of each of the plurality of lines;

selecting, for each of the plurality of lines, the segment having the largest sum;

identifying a smoothest region of the selected segments;

checking whether the smoothest region resembles a human upper body; and

extracting, as the candidate area, a [[the]] portion of the smoothest region that resembles a human head;

using one or more hierarchical verification levels to verify whether a human face is in the candidate area;

indicating that the candidate area includes a face if the one or more hierarchical verification levels verify that a human face is in the candidate area; and

using a plurality of cues to track each verified face in the content from frame to frame.

10.   (Original) A method as recited in claim 9, wherein determining whether there is motion comprises:

determining, for each of the plurality of pixels, whether a difference between an intensity value of the pixel in the frame and an intensity value of a corresponding pixel in one or more other frames exceeds a threshold value.

11.   (Original) A method as recited in claim 1, wherein the one or more hierarchical verification levels include a coarse level and a fine level, wherein the coarse level can verify whether the human face is in the candidate area faster but with less accuracy than the fine level.

12.   (Original) A method as recited in claim 1, wherein using one or more hierarchical verification levels comprises, as one of the levels of verification:

generating a color histogram of the candidate area;

generating an estimated color histogram of the candidate area based on previous frames;

determining a similarity value between the color histogram and the estimated color histogram; and

verifying that the candidate area includes a face if the similarity value is greater than a threshold value.

13. (Currently amended) A method as recited in claim 1, wherein indicating that the candidate area includes [[a]] the face comprises recording the candidate area in a tracking list.

14. (Original) A method as recited in claim 13, wherein recording the candidate area in the tracking list comprises accessing a record corresponding to the candidate area and resetting a time since last verification of the candidate.

15. (Original) A method as recited in claim 1, wherein the one or more hierarchical verification levels include a first level and a second level, and wherein using the one or more hierarchical verification levels to verify whether the human face is in the candidate area comprises:

checking whether, using the first level verification, the human face is verified as in the candidate area; and

using the second level verification only if the checking indicates that the human face is not verified as in the candidate area by the first level verification.

16. (Original) A method as recited in claim 1, wherein using one or more hierarchical verification levels comprises:

using a first verification process to determine whether the human head is in the candidate area; and

if the first verification process verifies that the human head is in the candidate area, then indicating the area includes a face, and otherwise using a second verification process to determine whether the human head is in the area.

17.   (Original)   A method as recited in claim 16, wherein the first verification process is faster but less accurate than the second verification process.

18.   (Original) A method as recited in claim 1, wherein the plurality of cues include foreground color, background color, edge intensity, motion, and audio.

19.   (Currently amended) A method ~~as recited in claim 1~~ comprising:

receiving a frame of content;

automatically detecting a candidate area for a new face region in the frame;

using one or more hierarchical verification levels to verify whether a human face is in the candidate area;

indicating that the candidate area includes a face if the one or more hierarchical verification levels verify that a human face is in the candidate area; and

using a plurality of cues to track each verified face in the content from frame to frame, wherein using the plurality of cues to track each verified face comprises, for each face:

predicting where a contour of the face will be;

encoding a smoothness constraint that penalizes roughness;

applying the smoothness constraint to a plurality of possible contour locations; and

Application No. 10/006,927

selecting the contour location having the smoothest contour as <u>a</u> [[the]] location of the face in the frame.

20.   (Original) A method as recited in claim 19, wherein the smoothness constraint includes contour smoothness.

21.   (Original) A method as recited in claim 19, wherein the smoothness constraint includes both contour smoothness and region smoothness.

22.   (Original) A method as recited in claim 19, wherein encoding the smoothness constraint comprises generating Hidden Markov Model (HMM) state transition probabilities.

23.   (Original) A method as recited in claim 19, wherein encoding the smoothness constraint comprises generating Joint Probability Data Association Filter (JPDAF) state transition probabilities.

24.   (Original) A method as recited in claim 19, wherein using the plurality of cues to track each verified face further comprises, for each face:

adapting the predicting for the face in subsequent frames to account for changing color distributions.

25.   (Original)   A method as recited in claim 19, wherein using the plurality of cues to track each verified face further comprises, for each face:

adapting the predicting for the face in subsequent frames based on one or more cues observed in the frame.

26.   (Currently amended)   A method ~~as recited in claim 1~~ comprising:

receiving a frame of content;

automatically detecting a candidate area for a new face region in the frame;

using one or more hierarchical verification levels to verify whether a human face is in the candidate area;

indicating that the candidate area includes a face if the one or more hierarchical verification levels verify that a human face is in the candidate area; and

using a plurality of cues to track each verified face in the content from frame to frame, wherein using the plurality of cues to track each verified face comprises, for each face:

accessing a set of one or more feature points of the face;

analyzing the frame to identify an area that includes the set of one or more feature points;

encoding a smoothness constraint that penalizes roughness;

applying the smoothness constraint to a plurality of possible contour locations; and

selecting the contour location having the smoothest contour as [[the]] a location of the face in the frame.

27. (Original) A method as recited in claim 1, wherein using the plurality of cues to track each verified face comprises concurrently tracking multiple possible locations for the face from frame to frame.

28. (Original) A method as recited in claim 27, further comprising using a multiple-hypothesis tracking technique to concurrently track the multiple possible locations.

29. (Original) A method as recited in claim 27, further comprising using a particle filter to concurrently track the multiple possible locations.

30. (Original) A method as recited in claim 27, further comprising using an unscented particle filter to concurrently track the multiple possible locations.

31-71. (Canceled).